Mean Field Theory of Dynamical Systems Driven by External Signals

Marc Massar[1] and Serge Massar[2]

[1] *370 Central Park West, Appt. 511, NY 10025 New York, USA*
[2] *Laboratoire d'Information Quantique, CP 225, Université libre de Bruxelles,*
Av. F. D. Roosevelt 50, B-1050 Bruxelles, Belgium
(Dated: October 2012)

Dynamical systems driven by strong external signals are ubiquituous in nature and engineering. Here we study "echo state networks", networks of a large number of randomly connected nodes, which represent a simple model of a neural network, and have important applications in machine learning. We develop a mean field theory of echo state networks. The dynamics of the network is captured by the evolution law, similar to a logistic map, for a single collective variable. When the network is driven by many independent external signals, this collective variable reaches a steady state. But when the network is driven by a single external signal, the collective variable is nonstationnary but can be characterised by its time averaged distribution. The predictions of the mean field theory, including the value of the largest Lyaponuov exponent, are compared with the numerical integration of the equations of motion.

PACS numbers:

Introduction. Our understanding of non linear dynamical systems and networks has made tremendous progress during the past decades. In most cases the autonomous dynamics is studied. The situation where the network is strongly driven by an external signal has so far been less investigated even though it arises in many different contexts in the natural and artificial world. Examples include networks of interacting chemicals (proteins, RNA) in a cell driven by unpredictable external chemical signals; networks of neurons driven by an external sensory input; artificial neural networks and their applications in machine learning; the response of population dynamics and ecological networks to changes in external conditions such as the weather; the responses of stock prices to economically significant news such as a company earnings, or unemployment numbers. In all these cases taking into account the external input is essential if one wants to understand correctly the dynamics, both because the external input is often large (it cannot be treated as a small perturbation), and because in some cases the systems itself has been selected according to its response to the fluctuating and unpredictable external variables.

The aim of the present work is to show, through the study of a specific but important example, how mean field techniques can provide a detailed understanding of dynamical networks strongly driven by an external signal. In the mean field approach the average feedback of the variables on themselves is taken into account through a self consistent equation, while the correlations between individual variables are neglected. The apparently extremely complicated dynamics of the network is thus reduced to much simpler evolution equations for a few collective variables. Previous applications of the mean field approach to dynamical systems (but without including an external input), and in particular neural networks, include e.g. [1–4]. For previous studies of dynamical systems in the presence of external signals (with however a quite different emphasiz than in the present work) see e.g. [5, 6].

The specific system we will consider is taken from the field of artificial neural networks. It consists of a network of randomly connected idealised neurons evolving in discrete time, and driven by an external time dependent signal, known in the machine learning community as an "echo state network" [7, 8], see also the continous time analog with no input studied in [4]. Such systems, when supplemented by a single linear output layer, fall within the class of "reservoir computers" [7–10] and currently hold records for several highly non trivial machine learning tasks such as time series prediction or some speech recognition benchmarks, see e.g. [11] for a review. Because of their importance in the machine learning community, it is highly desirable to better understand the dynamics of these systems. In addition they can serve as toy models for investigating the dynamics of neural networks, and more generally any recurrent dynamical systems, driven by external inputs.

An echo state network consists of a large number N of artificial neurons evolving in discrete time $t \in \mathbb{Z}$. We denote by $a_i(t)$ the "activation potential" of neuron i at time t. At time $t+1$, neuron i sends a signal to the other neurons with strength $x_i(t+1)$ given by

$$x_i(t+1) = f\left(a_i(t)\right) \ , \tag{1}$$

where the function f is taken to be a sigmoidal functions, i.e. f is odd, monotonously increasing, has finite limit for large a, and its first derivative $f'(a)$ decreases monotonously for positive a. By rescaling x and a we can redefine $f(a) \to \alpha f(\beta a)$. We choose the scales such that $f'(0) = 1$ and $\lim_{a\to\infty} f(a) = 1$. In the illustrative figures, we choose for f the hyperbolic tangent $f(a) = \tanh(a)$.

The update rule for the activation potentials is

$$a_i(t) = \sum_{j=1}^{N} w_{ij} x_j(t) + u_i s(t) \, , \qquad (2)$$

where w_{ij} is a time independent coupling matrix which gives the strength of the coupling of neuron j to neuron i, $s(t)$ is the time dependent external input, and u_i is a time independent vector which determines the strength with which the input is coupled to neuron i.

In echo state networks, the w_{ij} and u_i are chosen independently at random, except for global scaling factors $w_{ij} \to \alpha w_{ij}$, $u_i \to \beta u_i$. By adjusting these scaling factors and by using an optimised linear readout it is possible to obtain excellent performance on a variety of machine learning tasks. The general heuristic is that the factor α should be adjusted for the system to be at the threshold of chaos, whereupon the response of the neural network to the input is highly complex, but deterministic. Previous studies of the dynamics of echo state networks have aimed to understand how different dynamical regimes are related to changes in their information processing capability [12]. Most closely connected to the present work is [13] where, based on earlier work on feedforward networks [14], the information processing capability of echo state networks could be studied in the limit where the number N of neurons tends to infinity.

Mean Field approximation. The key insight behind the present work is to make the assumption that, at each time t, the $x_i(t)$ behave as independent identically distributed random variables which are also independent of the w_{ij} and the u_i. Then the term $\sum_{j=1}^{N} w_{ij} x_j(t)$ in eq. (2) is a sum of many identically distributed independent variables, and the law of large numbers tells us that this sum is distributed as a Gaussian, see fig. 1).

With this assumption we can compute the distribution of $a_i(t)$, and then using eq. (1) the distribution of $x_i(t+1)$. This analysis will yield a very simple one dimensional recurrence, similar to the logistic map, which captures the essence of the dynamics of the echo state network.

In more detail, we reason as follows. Because the function f is odd, the distribution from which are drawn the $x_i(t)$ has mean zero $\langle x_i(t) \rangle = 0$, where $\langle \rangle$ denotes ensemble average, i.e. average over the index i at fixed time t. We denote the variance of the $x_i(t)$ by $\langle x_i^2(t) \rangle = \sigma^2(t)$. Assuming that the w_{ij} are drawn independently at random from a distribution with mean zero $E[w_{ij}] = 0$ and variance $E[w_{ij}^2] = w^2$, and introducing the rescaled gain g as

$$g^2 = N w^2$$

we find that the term $\sum_{j=1}^{N} w_{ij} x_j(t) \sim N(0, g^2\sigma^2(t))$ has gaussian distribution. We now assume for simplicity that the u_i are drawn independently at random from a Gaussian distribution with zero mean. By rescaling $s(t)$ we
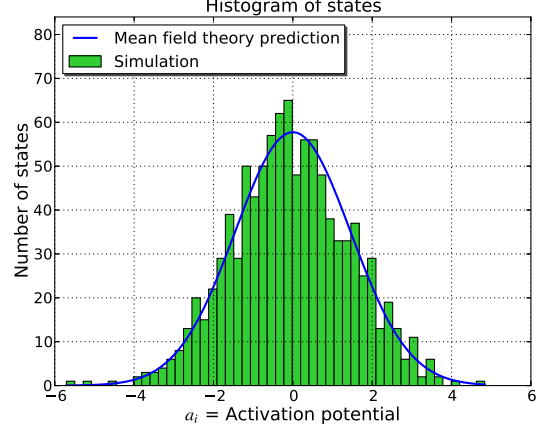

Figure 1: Distribution of the activation potential a_i. A reservoir with size $N = 1000$ and normalized gain of $g = 2$ and no source was run for 200 time steps, then the histogram of the activation potential was ploted in green. For comparaison a guassian with the variance predicted by the mean field theory was ploted in blue.

can, without loss of generality, assume that the gaussian has unit variance. Then the activation potential $a_i(t) = \sum_{j=1}^{N} w_{ij} x_j(t) + u_i s(t) \sim N(0, g^2\sigma^2(t) + s^2(t))$ also has a gaussian distribution. (If the u_i are drawn from a distribution other than Gaussian, then the distribution of the $a_i(t)$ can in principle be calculated, but the expressions will be more complicated). For future use we denote the variance of the $a_i(t)$ as $\Sigma^2(t) = g^2\sigma^2(t) + s(t)$. Finally, the distribution of x_i at time $t + 1$ is given by $x_i(t+1) \sim f(N(0, \Sigma^2(t)))$. We thus obtain a closed one dimensional recurrence for the variances of $x_i(t)$ and $a_i(t)$:

$$\begin{aligned} \Sigma^2(t) &= g^2\sigma^2(t) + s^2(t) \\ \sigma^2(t+1) &= F\left(\Sigma^2(t)\right) \end{aligned} \qquad (3)$$

where

$$\begin{aligned} F\left(\Sigma^2\right) &= \int da \, f^2(a) \, \frac{\exp\left[-\frac{a^2}{2\Sigma^2}\right]}{\sqrt{2\pi\Sigma^2}} \\ &= \int dx \, f^2(\Sigma x) \, \frac{\exp\left[-\frac{x^2}{2}\right]}{\sqrt{2\pi}} \end{aligned} \qquad (4)$$

From the properties of f, it folllows that $F\left(\Sigma^2\right)$ has the following properties: $F(0) = 0$, $F(+\infty) = 1$, $dF/d\Sigma^2 > 0$, $dF/d\Sigma^2(\Sigma^2 = 0) = 1$.

We first consider the case when there is no source $s^2 = 0$. When $g < 1$, there is a single stationnary solution $\sigma^2 = \Sigma^2 = 0$, corresponding to a quiescent system. $g = 1$ corresponds to a branching point. When $g > 1$, the stable stationnary solution is different from zero: $\sigma^2 > 0$, $\Sigma^2 > 0$. In the limit of infinite gain $g \to \infty$, we have $\sigma^2 \to 1$, $\Sigma^2 \to g^2$.
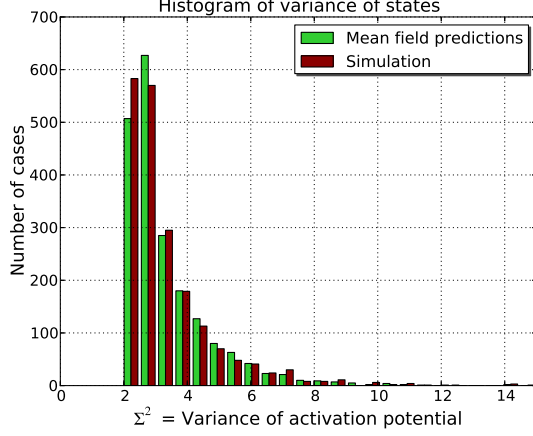
Figure 2: Histogram of the variance Σ^2 of the activation potential in the presence of source. A reservoir of size $N = 500$ with normalized gain $g = 2$ and a single i.i.d normal source with variance $\xi^2 = 1$ was run for 200 time steps to remove the influence of the initial state. Then the reservoir was run for 2,000 time steps, the variance of the activation potential $\Sigma^2(t)$ was collected, and its histogram ploted in red. The same procedure was done following the mean field theory predicion of eq. (2) and ploted in green.
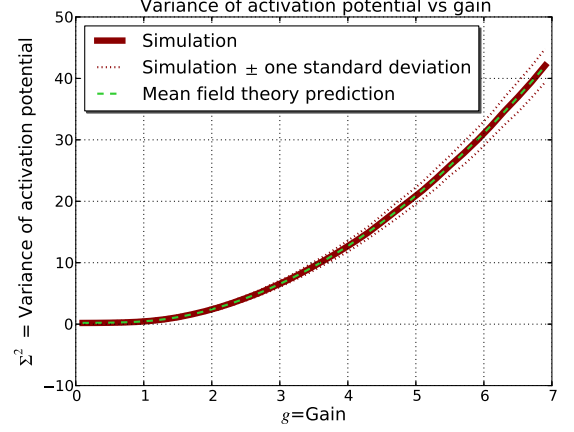

Figure 3: Variance of the activation potential Σ^2 as a function of normalised gain g. For each activation potential, a reservoir of size $N = 500$ with a single i.i.d. normal source with variance $\xi^2 = 0.2$ was run according to eqs. (1, 2) for 200 time steps to eliminate transitory effects, then the variance Σ^2 of the activation potential was recorded for 500 time steps. For each gain, the mean and standard deviation of Σ^2 was computed. The mean is ploted in red, and mean $\pm$ one standard deviation are plotted in doted red. The green dashed curve (superimposed on the red curve in the figure) corresponds to the predictions of the mean field theory computed with N independent sources according to eqs. (5, 6).

When the source $s(t)$ is non-zero, integrating the recurrence eq. (3) yields a distribution of values for σ^2 and Σ^2. In the Figures, we take for illustrative purposes the $s(t)$ to be independently drawn at each time t from the same probability distribution, that is we assume there are no temporal correlations between successive values of the source. For definitness we take the $s(t) \sim N(0, \xi^2)$ to be distributed according to a Gaussian with zero mean and variance ξ^2. Comparison of the mean field theory and the exact distribution for $\sigma^2(t)$ obtained by integrating the equations of motion is given in fig. 2.

It is instructive to compare the above situation with the case where the echo state network is driven by many independent sources. In this case one replaces eq. (2) by $a_i(t) = \sum_{j=1}^{N} w_{ij}x_j(t) + v_i(t)$ with $v_i(t)$ taken to be independently drawn from the Gaussian distribution $v_i(t) \sim N(0, v^2)$. In this case the recurrence (3) becomes

$$\Sigma^2(t) = g^2\sigma^2(t) + v^2 \qquad (5)$$
$$\sigma^2(t+1) = F\left(\Sigma^2(t)\right) \qquad (6)$$

This recurrence admits a stationnary solution. When $v^2 > 0$, the stationnary solution is always different from zero: $\sigma^2 > 0$, $\Sigma^2 > 0$. For very small gain $g \to 0$ we have $\Sigma^2 = v^2$, $\sigma^2 = F(v^2)$. In fig. 3 we show that the variance of an echo state network driven by a single source is close to the stationary variance computed according to eqs. (5, 6): the time average of the variance is identical to the stationary solution, and the standard deviation is small.

Note that in the limit of infinite gain $g \to \infty$, we have $\sigma^2 \to 1$, $\Sigma^2 \to g^2$, independently of the source.

In the figures we took $f(a) = \tanh(a)$ whereupon the integral eq. (4) yielding the sationnary solution F had to be carried out numerically. This case is of interest as it is the one most often used in previous works. It is however interesting to note that the function $F(\Sigma^2)$ can be computed analytically in two cases. The first does not correspond to a sigmoid function f, but it is of interest as it has been used in a recent experimental implementation of reservoir computing [15, 16]: if $f(a) = \sqrt{2}\sin(\frac{a}{\sqrt{2}})$, then $F(\Sigma^2) = 1 - \exp\left(\Sigma^2\right)$. The second case (obtained by first evaluating $\frac{dF(\Sigma^2)}{d\Sigma^2}$, see [14]) corresponds to a sigmoid function f: if $f(a) = \text{erf}(\frac{\sqrt{\pi}}{\sqrt{2}}a)$, then $F(\Sigma^2) = -1 + \frac{4}{\pi}\arctan\left(\sqrt{1 + 2\Sigma^2}\right)$.

Stability. The mean field theory also allows a derivation of the Lyapunov exponents of the echo state network. Consider two neighbouring solutions $a_i(t)$ and $a_i'(t)$, with joint Gaussian distribution: $\frac{a_i(t)+a_i'(t)}{2} \sim N(0, \Sigma^2(t))$, $a_i(t) - a_i'(t) \sim N(0, \delta^2(t))$. We take $\delta^2(t)$ small and wish to compute how it evolves with time. We have $a_i(t+1) - a_i'(t+1) = \sum_j w_{ij}f'\left(\frac{a_i(t)+a_i'(t)}{2}\right)(a_i(t) - a_i'(t)) + O\left((a-a')^2\right)$ from which we derive that

$$\delta^2(t+1) = \delta^2(t)\Lambda(\Sigma^2(t)) + O\left(\delta^3(t)\right)$$

$$\Lambda(\Sigma^2) = g^2 \int da\ f'^2(a)\ \frac{\exp\left[-\frac{a^2}{2\Sigma^2}\right]}{\sqrt{2\pi\Sigma^2}}\ . \qquad (7)$$
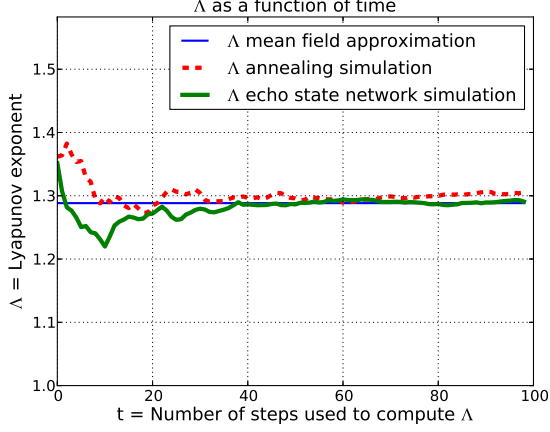
Figure 4: Convergence of Lyapunov exponent estimate. A single reservoir of size $N = 500$, normalized gain $g = 2$, and source variance $\xi^2 = 0.2$ was run for 200 steps. Then the states were perturbed with i.i.d. gaussians with standard deviation 10^{-10} and the size of the pertubation was recored for 100 time steps. The lyapunov exponent Λ was computed for the different lags: $\sqrt[t]{\delta^2(t)/\delta^2(0)}$. The green line is the result of a simulation done according to eqs. (1, 2), the doted red line is the result of a simulation done in the annealed approximation (see conclusion), the straight blue line is the mean field theory prediction computed from (7).

See fig. 4 for a comparison between Lyapunov exponents computed from the exact equations of motion and from the mean field theory.

From this expression two interesting properties immediately follow. First, in the absence of source the quiescent stationnary solution $\sigma^2(t) = \Sigma^2(t) = 0$ has largest Lyapunov exponents smaller than 1 for $g < 1$, and largest Lyapunov exponents larger than 1 for $g > 1$. (This follows from the fact that the integral in eq. (7) equals 1 in the limit $\Sigma^2 \to 0$, since $f'(0) = 1$). Thus, in the absence of source, $g = 1$ is the threshold for chaos in the dynamical system.

Second, for sigmoidal functions f, we find the property (well known to those who use echo state networks for machine learning tasks) that increasing the strength of source term ξ^2 stablizes the system. Indeed when ξ^2 increase, Σ^2 also increases. For sigmoidal functions $f'(a)$ is a decreasing function of $|a|$, and hence the integral in eq. (7) decreases when Σ^2 increases. See fig. 5 for illustrations of this prediction.

Conclusion. In the present work we have shown that mean field theory can be applied to large random networks driven by external signals. The agreement, at least in the case of echo state networks, with the exact integration of the equations of motion is remarkably good.

The mean field theory itself is closely related to the *annealed approximation* wherein the $w_{ij}(t)$ and the $u_i(t)$ are redrawn independently at random at each time t from the same distribution, i.e. the coupling coefficients become
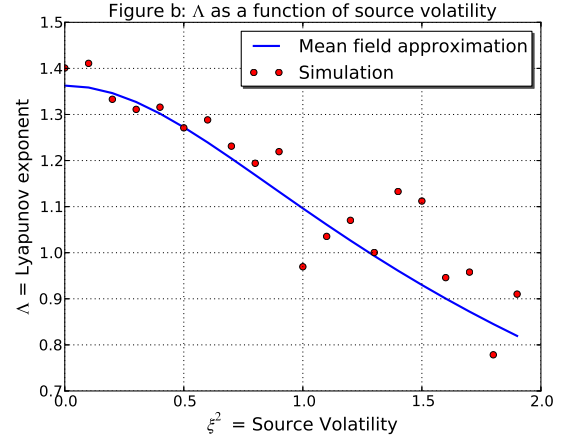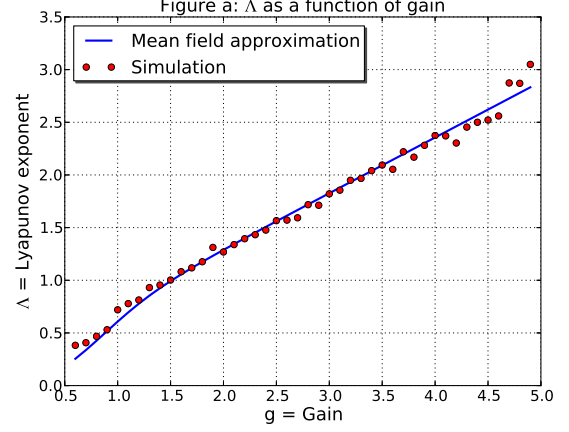

Figure 5: Lyapunov exponent as a function of gain (panel a) and Lyapunov exponent as a function of source volatility (panel b). An $N = 500$, $g = 2$ reservoir was run 200 time steps, than perturbed by i.i.d. gaussians with standard deviation 10^{-12} and run for 20 time steps. The lyapunov exponent was computed based on how much the pertubation changed, and ploted in red. The mean field theory prediction is in blue. In panel a the source variance is $\xi^2 = 0.2$; in panel b the gain is $g = 2$.

time dependent. We expect the mean field theory to be exact in the large N limit of the annealed approximation.

Compared with the case when there is no external signal, we recover a number of features which are well known empirically to people working with echo state networks, but which have not been derived analytically before. In particular we find that if in the absence of external signal the system has a trivial stable state (corresponding in our analysis to the case $g < 1$), then in the presence of external signal the dynamics becomes non trivial. We also find that the presence of the external signal tends to stabilize the system (i.e. Lyapunov exponents which decrease when the external signal increases).

We also compare the cases where the dynamical system is driven by a many independent input signals, and by a single input signal. We find a qualitative difference.

Namely in the first case the mean field theory predicts that the collective variables take on stationnary values, whereas in the second case the mean field theory predicts their statistical distribution.

The present work should provide the basis for further in depth understanding of the highly important case in practice of dynamical systems driven by time dependent external signals.

-----

[1] D. J. Amit and M. Tsodyks, Network **2**, 259, 1991

[2] A. Treves, Network **4**, 259, 1993

[3] M. Mattia and P. Del Giudice, Phys. Rev. E **66**, 051917, 2002

[4] H. Sompolinsky, A. Crisanti, H. J. Sommers, Phys. Rev. Lett. **61**, pp. 259-262 (1988)

[5] B. Lindnera, J. Garcia-Ojalvob, A. Neimand, L. Schimansky-Geier, Phys. Reports **392**, 321–424 (2004)

[6] F. Sagues, J. M. Sancho, J. Garcia-Ojalvo, Rev. Mod. Phys. **79**, 829–882 (2007)

[7] H. Jaeger, Fraunhofer Institute for Autonomous Intelligent Systems, Technical report: GMD Report 148, 2001.

[8] H. Jaeger and H. Haas, Science **78**, pp. 78-80, 2004

[9] W. Maas, T. Natschlager, H. Markram, Neural Computation **14**, 2531-2560, 2002.

[10] D. Verstraeten, B. Schrauwen, M. D. Haene, D. Stroobandt, Neural Networks **20**, pp. 391-403, 2007.

[11] M. Lukocevičius, H.Jaeger, Computer Science Review 3, pp. 127-149, 2009.

[12] B. Schrauwen, L. Büsing, and R. Legenstein, Advances in Neural Information Processing Systems **21**, 1425. (2009)

[13] M. Hermans, B. Schrauwen, Neural Computation **24**, pp. 104-133, 2012

[14] C. K. A. Williams, Neural Computation **10**, pp. 1203-1216, (1998)

[15] Y. Paquot et al., Scientific Reports **2** 287, 2012.

[16] L. Larger et al., Optics Express **20** 3241, 2012.